DECOMPOSITION IN CONVEX PROGRAMMING AND OPTIMAL CONTROL PROBLEMS

BERND LUDERER* and BERND SCHWARTZ**
* Technische Universität Karl-Marx-Stadt, Sektion Mathematik, DDR – 9010 Karl-Marx-Stadt, PSF 964
** Technische Hochschule Ilmenau, Sektion MARÖK, DDR – 6300 Ilmenau, PSF 327

(Received March 12, 1984)

0. Introduction

In recent years there has been an increasing interest in studying decomposition methods, which replace a large-scale problem by a sequence of smaller problems (see e.g. [1], [2], [5], [11]). This interest comes, on the one hand, from many practical applications, in which large-scale programming problems of a certain structure normally arise, on the other hand, from the fact that structured problems of high dimension as pure mathematical problems, also occur inner mathematically, for instance, in solving optimal control problems numerically, say by discretization. As a rule, decomposition algorithms work effectively only if the problem to be dealt with possesses some structure. Moreover, if the considered problem is of a dimension such that it is impossible to solve it even on a large computer, the use of decomposition techniques may be the unique way of solving the problem.

Another interesting class of problems is related to the control of hierarchical systems, where often so-called multilevel methods are used.

The aim of this paper is to contribute to a unified treatment of both classes of problems. For this purpose we consider a structured convex programming problem as well as an optimal control problem and describe decomposition algorithms (two-level methods) for them, using feasible methods. They are based on the notion of the subdifferential of a certain function and enlargements of this set (ε-subdifferential and other modifications).

1. Statement and substitute problems

In the present paper two examples of structured problems are to be considered. The first comes from convex programming:

$$(1) \qquad \sum_{i=1}^{N} f_i(x_i) \to \inf; \ \sum_{i=1}^{N} g_i(x_i) \leq 0, \quad h_i(x_i) \leq 0, \quad i = 1, \ldots, N.$$

We assume that the set of optimal solutions of (1) is non-empty and compact, and the functions $f_i : \mathbf{R}^{n_i} \to \mathbf{R}$, $g_i : \mathbf{R}^{n_i} \to \mathbf{R}^m$, $h_i : \mathbf{R}^{n_i} \to \mathbf{R}^{s_i}$, $i = 1, \ldots, N$

are convex and differentiable on the neighbourhoods of the sets $X_{i0} = \{x_i | h_i(x_i) \leq 0\}$, where the sets X_{i0}, $i = 1, \ldots, N$ fulfil the Slater condition.

As can be shown, the following substitute problem is equivalent to (1):

$$\Phi(y) = \Phi(y_1, \ldots, y_N) = \sum_{i=1}^{N} \varphi_i(y_i) \to \inf; \sum_{i=1}^{N} y_i = 0, \tag{2}$$

where

$$\varphi_i(y_i) = \inf \{f_i(x_i) | g_i(x_i) \leq y_i,\ h_i(x_i) \leq 0\}, \tag{3i}$$

$i = 1, \ldots, N$. The function Φ is defined implicitly. This means that for a given y we determine the value $\Phi(y)$ by the help of the optimal solutions of the arising subproblems (3i). Then we must find the optimal $\hat{y}$ subject to the equality restriction.

The second problem under consideration is a structured optimal control problem connected with a system consisting of interconnected subsystems:

$$\begin{cases} \mathcal{J}(x(\cdot), u(\cdot), z(\cdot)) = \int_{t_0}^{t_1} \sum_{i=1}^{N} f_i(t, x_i, u_i, z_i) dt \to \inf; \\ \dot{x} = A_i x_i + B_i u_i + C_i z_i, \quad x_i(t_0) = x_{i0}, \quad y_i = D_i x_i + E_i u_i, \\ z_i = F_i y = \sum_{j=1}^{N} F_{ij} y_j, \quad u_i \in U_i, \quad i = 1, \ldots, N. \end{cases} \tag{4}$$

Here we assume that t_0, t_1 and $x_{i0} \in \mathbf{R}^{n_i}$ are fixed quantities, the states $x_i(t)$ are elements of $W_{2,1}^{n_i}([t_0, t_1])$ and the controls $u_i(t)$, subsystem inputs $z_i(t)$ and outputs $y_i(t)$ are elements of the spaces $L_2^{r_i}$, $L_2^{m_i}$ and $L_2^{s_i}$ respectively. Furthermore, let A_i, B_i, C_i, D_i, E_i, F_{ij} be matrices of appropriate dimensions (may be depending on time t). Finally, assume that $U_i \subset \mathbf{R}^{r_i}$ are convex sets and f_i are convex functions, either continuously differentiable in (x_i, u_i, z_i) or additive.

Again we are able to formulate an equivalent substitute problem to (4) in the following manner:

$$\Psi(y(\cdot)) = \Psi(y_1(\cdot), \ldots, y_N(\cdot)) = \sum_{i=1}^{N} \psi_i(y(\cdot)) \to \inf, \tag{5}$$

where

$$\begin{aligned} \psi_i(y(\cdot)) = \inf \Big\{ \int_{t_0}^{t_1} f_i(t, x_i, u_i, z_i) dt \,|\, \dot{x}_i = A_i x_i + B_i u_i + C_i z_i, \\ x_i(t_0) = x_{i0},\ D_i x_i + E_i u_i = y_i,\ z_i = F_i y,\ u_i \in U_i \Big\}, \quad i = 1, \ldots, N. \end{aligned} \tag{6i}$$

Thus, for given $y(\cdot)$ the value $\Psi(y(\cdot))$ is determined as the optimal solution of the optimal control problem (4). Then the optimal $\hat{y}(\cdot)$ is evaluated. Because of the structure of the problem, (5) falls into N subproblems (6i) if $y(\cdot)$ is fixed.

Of course, the vectors y and vector functions $y(\cdot)$ in (2) and (5), respectively, must belong to the set $\mathcal{Y}$ of elements such that the subproblems (3i) and (6i) on the lower level are solvable.

Our aim is to solve problems (2) and (5) instead of (1) and (4). According to the assumptions, the functions $\Phi(y)$ and $\Psi(y(\cdot))$ are both convex. However, it should be mentioned that these functions are nondifferentiable in general. Further, we emphasize that to evaluate a value of Φ or Ψ means to solve N subproblems (3i) or (6i).

Because of the convexity of Φ and Ψ, algorithms of convex programming for solving (2) and (5) can be used (e.g. feasible direction or subgradient methods), which work as a rule on the basis of the subdifferential of Φ and Ψ or some enlargements of this set. The methods considered below are of two-level type and decompose the original problem into a number of smaller subproblems, which are to be solved several times. They belong to the class of feasible (or primal) methods. For the optimal control problem (4) this presupposes that there are at least as many controls as interactions.

2. Subdifferential representations

Let f be a proper convex function on the Banach space X, and let $\varepsilon \geq 0$. Then the set $\partial_\varepsilon f(x_0) = \{x^* \in X^* | \langle x^*, x - x_0 \rangle \leq f(x) - f(x_0) + \varepsilon, \ \forall x \in X\}$ is said to be the ε-subdifferential of f at x_0. The set $\partial_0 f(x_0) = \partial f(x_0)$ is simply called the subdifferential.

a) In the following, let $\bar{x}_i$ be an optimal solution of problem (3i) for given $\bar{y}_i$. Under the assumptions formulated above the subdifferential of the function Φ at the point $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_N)$ has the form

$$\text{(7)} \qquad \partial\Phi(\bar{y}) = \mathop{\times}_{i=1}^{N} \partial\varphi_i(\bar{y}_i) = \{(y_1^*, \ldots, y_N^*) | y_i^* \in \partial\varphi_i(\bar{y}_i), i = 1, \ldots, N\},$$

with

$$\partial\varphi_i(\bar{y}_i) = \{y_i^* \in \mathbf{R}^m | y_i^* \leq 0, \ \exists \bar{\lambda}_i \in \mathbf{R}^{s_i}: \bar{\lambda}_i \geq 0, \ f_i'(\bar{x}_i) - g_i'^T(\bar{x}_i) y_i^* +$$
$$+ h_i'^T(\bar{x}_i)\bar{\lambda}_i = 0, \ \langle \bar{\lambda}_i, h_i(\bar{x}_i) \rangle + \langle y_i^*, \bar{y}_i - g_i(\bar{x}_i) \rangle \geq 0\},$$

where $g'(x_i)$ denotes the matrix of partial derivatives of g_i. The proof of this statement may be found in [10]. Here the vectors y_i^* and $\bar{\lambda}_i$ are Lagrange multipliers corresponding to the restrictions g_i and h_i of (3i). According to the representation described above the subdifferentials $\partial\varphi_i$ and, therefore, also $\partial\Phi$ can be generated when arbitrary solutions x_i and corresponding Lagrange multipliers y_i^* and $\bar{\lambda}_i$ of the subproblems are known.

b) In the special case of problem (1) when all functions are linear, i.e. if $f_i(x_i) = \langle c_i, x_i \rangle$, $g_i(x_i) = A_i x_i$, $h_i(x_i) = B_i x_i - b_i$, the following representation holds:

$$\partial_\varepsilon\varphi_i(\bar{y}_i) = \{y_i^* \in \mathbf{R}^m | -\langle y_i^*, \bar{y}_i \rangle + \langle b_i, \bar{\lambda}_i \rangle \leq -\varphi_i(\bar{y}_i) + \varepsilon,$$
$$y_i^* A_i - \bar{\lambda}_i B_i = c_i, \ y_i^* \leq 0, \ \bar{\lambda}_i \geq 0\},$$

(8) $\partial_\varepsilon \Phi(\bar{y}) = \left\{(y_1^*, \ldots, y_N^*) \mid y_i^* \in \partial_{\varepsilon_i}\varphi(\bar{y}_i),\ \sum_{i=1}^{N} \varepsilon_i = \varepsilon,\quad \varepsilon_i \geq 0,\ i = 1, \ldots, N\right\}.$

c) Let the pair $(\bar{x}_i(t), \bar{u}_i(t))$ be a solution of the subproblem (6i), $i = 1, \ldots, N$, for given $\bar{y}(\cdot)$, and let $\bar{p}_i(t)$ and $\bar{q}_i(t)$ be corresponding Lagrange multipliers of the first two restrictions of the i-th subproblem (6i). Then (cf. [7])

(9) $$\partial\Psi(\bar{y}(t)) = \sum_{i=1}^{N} \partial\psi_i(\bar{y}(t)),$$

where

$$\partial\psi_i(\bar{y}(t)) = \{y_i^*(t) \mid y_i^*(t) = F_i^T f_{iz_i}(t, \bar{x}_i(t), \bar{u}_i(t), F_i\bar{y}(t)) - F_i^T C_i^T \bar{p}_i(t) - (0, \ldots, \bar{q}_i(t), \ldots, 0)^T\}.$$

An analogous representation can be obtained in the case if the convex (possibly nondifferentiable) functions f_i are additive.

3. Method of feasible directions for solving (2)

First of all, we rewrite (2) into a free minimum problem:

(10) $$\hat{\Phi}(y) = \Phi(y) + \delta(y \mid Y),$$

where $Y = \left\{y \mid \sum_{i=1}^{N} y_i = 0\right\}$ and δ is the indicator function. For solving (10) we study a new method of feasible directions for nondifferentiable functions. As is well-known there arise two main problems related with such a method: to find for a given point y^k a direction r^k such that the directional derivative $\hat{\Phi}'(y^k; r^k) < 0$, and to determine the optimal step size t_k. The main task on which we focus our attention is the determination of an appropriate direction r^k.

For this purpose let $\{\varepsilon_l\}_{l=1}^{\infty} \downarrow 0$ be a monotonically decreasing sequence and $\{g_l(y, r)\}_{l=1}^{\infty}$ a sequence of functions, closely related to $\hat{\Phi}'(y; r)$ and fulfilling certain assumptions (see [10]). Then the principal procedure may be described as follows:

0°. Set $k := 0$, $l := 0$, choose $x_0 \in \operatorname{dom} f$.

1°. Determine r^k from $g_l(y^k, r^k) = \min\{g_l(y^k, r) \mid \|r\| \leq 1\}$. If $g_l(y^k, r^k) < -\varepsilon_l$, then go to 2° else to 4°.

2°. Find t_k from $\hat{\Phi}(y^k + t_k r^k) = \min\{\hat{\Phi}(y^k + t r^k) \mid t \geq 0\}$.

3°. Update $y^{k+1} := y^k + t_k r^k$, set $k := k+1$, go to 1°.

4°. $l := l+1$, i.e. diminish ε_l, go to 1°.

Remark. In Step 1° it is not necessary to evaluate just the direction r^k which yields the exact minimum of g_l, but we may choose an arbitrary direction r^k such that $g_l(y^k, r^k) < -\varepsilon_l$.

It is natural to use the function

$$g_l(y, r) = \sup \{\langle r, w\rangle \mid w \in P_l(y)\}, \tag{11}$$

where P_l is a point-to-set mapping with $P_l(y) \supset \partial\hat{\Phi}(y)$. How can a suitable set $P_l(y)$ be chosen? Let us begin with the linear case of problem (1):

$$P_l(y) = \{w = (w_1, \ldots, w_N) \mid w_i = y_i^* + \mu,\ y_i^* \in \partial_{\varepsilon_l} \varphi_i(y_l)\ \ \forall\, i, \mu \in \mathbf{R}^m\}.$$

Owing to this definition of $P_l(y)$, the inclusion

$$\partial_{\varepsilon_l}\hat{\Phi}(y) \subset P_l(y) \subset \partial_{N\varepsilon_l}\hat{\Phi}(y) \tag{12}$$

holds. In this case the following programming problem may be considered as the problem of finding a feasible direction: find a direction $r_l = (r_{1l}, \ldots, r_{Nl})$ with $\|r_l\| \leq 1$, $\sum_{i=1}^{N} r_{il} = 0$ and $h(r_l) \overset{\text{def}}{=} \sup \left\{\sum_{i=1}^{N} \langle r_{il}, w_i\rangle \mid w \in P_l(y)\right\} < 0$. This problem can be solved by determining a hyperplane separating the origin and the set $P_l(y)$ with regard to the representation (8). Several methods implementable on a computer are described in [10].

According to (12) the following can be said. If for all r_l with $\sum_{i=1}^{N} r_{il} = 0$ the inequality $h(r_l) \geq 0$ holds (i.e. if $0 \in P_l(y)$), then $\hat{\Phi}(y) \leq \min \hat{\Phi} + N\varepsilon_l$, which means that y is a $N\varepsilon$-optimal solution. For $\bar{r}_l$ such that $\sum_{i=1}^{N} \bar{r}_{il} = 0$ and $h(\bar{r}_l) < 0$ the relation $\min \{\hat{\Phi}(y + t\bar{r}_l) \mid t \geq 0\} \leq \hat{\Phi}(y) - \varepsilon_l$ is valid, i.e., if $0 \notin P_l(y)$ and, therefore, $0 \notin \partial_{\varepsilon_l}\hat{\Phi}(y)$, an improvement of at least ε_l is ensured. The situation is similar in the convex case of problem (1), too. For details see [10].

4. A subgradient algorithm for solving (5)

We note that in principle it is possible to apply the above algorithm also to the optimal control problem (4). However, this has not yet been studied very well up to now. Here we consider a special case when $U_i = \mathbf{R}^{r_i}$, i.e., if no constraints on controls are present. We shall describe a subgradient method for solving (5) which is based on the representation of the subdifferential $\partial\Psi$ given above.

As was mentioned in Section 1, problem (5) falls into N subproblems if we fix an admissible vector $\bar{y}(t) = (\bar{y}_1(t), \ldots, \bar{y}_N(t))$ and, therefore, also fix $\bar{z}(t) = (\bar{z}_1(t), \ldots, \bar{z}_N(t))$ with $\bar{z}_i(t) = F_i\bar{y}(t)$. It means, for given $\bar{y}(t)$, we have to solve the N problems

$$\text{(13i)}\quad \begin{cases} \mathcal{J}_i(x_i(\cdot), u_i(\cdot)) = \int_{t_0}^{t_1} f_i(t, x_i, u_i, \bar{z}_i)\,dt \to \inf; \\ \dot{x}_i = A_i x_i + B_i u_i + C_i \bar{z}, \quad x_i(t_0) = x_{i0}, \quad D_i x_i + E_i u_i = \bar{y}_i. \end{cases}$$

Solving (5) means to determin a $\hat{y}(t)$ such that the optimal solutions of the subproblems (13i), $i = 1, \ldots, N$ yield an optimal solution of the overall problem (4). For this purpose we want to apply a subgradient algorithm. According to the relation (9) we must, for given $y^{(k)}(t)$, determine $(x_i^{(k)}(t), u_i^{(k)}(t), p_i^{(k)}(t), q_i^{(k)}(t))$ from the system

$$\text{(14i)} \quad \left. \begin{aligned} &\dot{x}_i - A_i x_i - B_i u_i = C_i F_i y^{(k)}, \ x_i(t_0) = x_{i0}, \\ &D_i x_i + E_i u_i = y_i^{(k)}, \\ &\dot{p}_i - A_i^T p_i + D_i^T q_i + f_{ix_i}(t, x_i, u_i, F_i y^{(k)}) = 0, \ p_i(t_1) = 0, \\ &-B_i^T p_i + E_i^T q_i + f_{iu_i}(t, x_i, F_i y^{(k)}) = 0, \end{aligned} \right\}$$

$i = 1, \ldots, N$, by any available method. Then we evaluate the required subgradient of $\Psi(y^{(k)}(\cdot))$ by the help of the formula (9), where $(\bar{x}_i, \bar{u}_i, \bar{p}_i, \bar{q}_i)$ are replaced by $(x_i^{(k)}, u_i^{(k)}, p_i^{(k)}, q_i^{(k)})$. Since there are no restrictions on u, the systems (14i) are, under mild assumptions, solvable for every $y(t)$, if the problems (13i) possess a minimum. Consequently, in this case, the set $\mathcal{Y}$ is the whole space and no projection is needed. Thus, it is possible to use the following subgradient method:

0°. Choose numbers $\gamma_k > 0$, $\sum_{k=0}^{\infty} \gamma_k = \infty$, $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$ and an admissible function $y^{(0)}(t)$; set $k := 0$.
1°. Solve (14i) for given $y^{(k)}(t)$. Determine $y^{(k)*}(t)$ according to (9). If $y^{(k)*}(t) \equiv 0$ stop: $y^{(k)}(t)$ is optimal.
2°. Evaluate $y^{(k+1)}(t) := y^{(k)}(t) - \gamma_k y^{(k)*}(t)$, go to 1°.

The convergence of this scheme has been proved in [9].

Remarks. 1. If there are restrictions on the controls (e.g. inclusions $u_i \in U_i$) or mixed equality or inequality relations for the subsystems $G(t, x_i, u_i) \leq 0$, the subgradient method in the form described above cannot be used, in view of the fact that the new vector function $y^{(k+1)}(t)$ must remain admissible, which is in general not true. Therefore, other procedures such as, e.g., the method of feasible directions considered above, have to be applied. Global restrictions can also be regarded if the described algorithm is modified in an appropriate manner (cf. also [4]).

2. Analogous statements (subdifferential representations) can be made and corresponding algorithms may be derived using the dual approach – the approach of goal coordination (cf. [8], [11]) where problems similar to (4) are discussed).

3. For a linear-quadratic optimal control problem similar to the one considered here (but without interactions) a subgradient algorithm was studied in [3], [6].

5. Final remarks

Note that the approach described above, applies, in principle, also to nonconvex problems. However, in this case a perfectly developed theory does not exist until now.

During the period of the last ten years the optimization team of the Department of Mathematics of Technical University Karl-Marx-Stadt (GDR) has intensively studied large-scale structured optimization problems (linear, convex, optimal control and other problems). At present, two resource allocation type methods for very large linear programming problems have been developed and implemented for the ES 1040 computer (up to 4000 restrictions in each subproblem, the number of variables is virtually unlimited). They use the programming system OPSI (OPtimization by SImplex method) and modify it. Problems of this kind and scope arise e.g. in production planning in industry and agriculture.

REFERENCES

[1] *Beer K.:* Lösung großer linearer Optimierungsaufgaben. VEB Deutscher Verlag d. Wissenschaften, Berlin, 1977.

[2] *Findeisen W.:* Hierarchische Steuerungssysteme (Transl. from Polish). VEB Verlag Technik, Berlin, 1977.

[3] *Franke D.* und *Luderer B.:* Realisierung eines Algorithmus zur Lösung von Aufgaben der optimalen Steuerung. *Wiss. Zeitschrift TH Karl-Marx-Stadt* **7** (1979), 853–858.

[4] *Käschel J.* und *Luderer B.:* Verfahren der zulässigen Richtungen in primalen Dekompositionsmethoden. *Wiss. Berichte TH Leipzig* **20** (1982), 43–48.

[5] *Lasdon L. S.:* Optimization Theory for Large Systems. The Macmillan Co., New York, 1970.

[6] *Luderer B.:* Lösung strukturierter Aufgaben der optimalen Steuerung mittels Zerlegungsverfahren. *Math. Operationsforsch. Statist., Ser. Optimization* **11** (4) (1980), 593–603.

[7] *Luderer B.:* Subdifferential representation of implicitly defined functions. *Math. Operationsforsch. Statist., Ser. Optimization* **13** (2) (1982), 179–186.

[8] *Pearson J. D.:* Dynamic decomposition techniques, ch. 4 In: Optimization Methods for Large-Scale Systems with Applications (ed. D. A. Wismer). McGraw-Hill, New York, 1971.

[9] *Поляк Б. Т.*: Один общий метод решения экстремальных задач. *ДАН СССР 174* (1) (1967), 33–36.

[10] *Käschel J.* und *Schwartz B.:* Verfahren der zulässigen Richtungen für Optimierungsaufgaben mit nichtdifferenzierbaren Funktionen und Dekomposition. *Wiss. Schriftenreihe TH Karl-Marx-Stadt* **11** (1981).

[11] *Singh M. G.:* Dynamical Hierarchical Control. North-Holland, Amsterdam/New York/Oxford, 1980.